

# The Distribution of the Asymptotic Number of Citations to Sets of Publications by a Researcher or From an Academic Department Are Consistent With a Discrete Lognormal Model

João A. G. Moreira<sup>1</sup>, Xiao Han T. Zeng<sup>1</sup>, and Luís A. Nunes Amaral<sup>1,2,3,4</sup>

<sup>1</sup>Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL, USA

<sup>2</sup>Department of Physics and Astronomy, Northwestern University, Evanston, IL, USA

<sup>3</sup>Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL, USA

<sup>4</sup>Howard Hughes Medical Institute, Northwestern University, Evanston, IL, USA

## Abstract

How to quantify the impact of a researcher’s or an institution’s body of work is a matter of increasing importance to scientists, funding agencies, and hiring committees. The use of bibliometric indicators, such as the  $h$ -index or the Journal Impact Factor, have become widespread despite their known limitations. We argue that most existing bibliometric indicators are inconsistent, biased, and, worst of all, susceptible to manipulation. Here, we pursue a principled approach to the development of an indicator to quantify the scientific impact of both individual researchers and research institutions grounded on the functional form of the distribution of the asymptotic number of citations. We validate our approach using the publication records of 1,283 researchers from seven scientific and engineering disciplines and the chemistry departments at the 106 U.S. research institutions classified as “very high research activity”. Our approach has three distinct advantages. First, it accurately captures the overall scientific impact of researchers at all career stages, as measured by asymptotic citation counts. Second, unlike other measures, our indicator is resistant to manipulation and rewards publication quality over quantity. Third, our approach captures the time-evolution of the scientific impact of research institutions.

## Introduction

The explosive growth in the number of scientific journals and publications has outstripped researchers’ ability to evaluate them [1]. To choose what to browse, read, or cite from a huge and growing collection of scientific literature is a challenging task for researchers in nearly all areas of Science and Technology. In order to search for worthwhile publications, researchers are thus relying more and more on heuristic proxies – such as author and journal reputations – that signal publication quality.

The introduction of the *Science Citation Index* (SCI) in 1963 [2] and the establishment of bibliographic databases spurred the development of bibliometric measures for quantifying the impact of individual researchers, journals, and institutions. Various bibliometric indicators have been proposed as measures of impact, including such notorious examples as the Journal Impact Factor and the  $h$ -index [3, 4]. However, several studies revealed that these measures can be inconsistent, biased, and, worst of all, susceptible to manipulation [5–15]. For example, the limitations of the popular  $h$ -index include its dependence on discipline and on career length [16].

In recent years, researchers have proposed a veritable alphabet soup of “new” metrics – the  $g$ -index [17], the  $R$ -index [18], the  $ch$ -index [19], among others – most of which are *ad-hoc* heuristics, lacking insight about why or how scientific publications accumulate citations.

The onslaught of dubious indicators based on citation counts has spurred a backlash and the introduction of so-called “altmetric” indicators of scientific performance. These new indicators completely disregard citations, considering instead such quantities as number of article downloads or article views, and number of “shares” on diverse social platforms [20–22]. Unfortunately, new research is showing that altmetrics are likely to reflect popularity rather than impact, that they have incomplete coverage of the scientific disciplines [23, 24], and that they are *extremely susceptible to manipulation*. For example, inflating the findings of a publication in the abstract can lead to misleading press reports [25], and journals’ electronic interfaces can be designed to inflate article views and/or downloads [26].

Citations are the currency of scientific research. In theory, they are used by researchers to recognize prior work that was crucial to the study being reported. However, citations are also used to make the research message more persuasive, to refute previous work, or to align with a given field [27]. To complicate matters further, the various scientific disciplines differ in their citation practices [28]. Yet, despite their limitations, citations from articles published in reputable journals remain the most significant quantity with which to build indicators of scientific impact [12].

It behooves us to develop a measure that is based on a thorough understanding of the citation accumulation process and also grounded on a rigorous statistical validation. Some researchers have taken some steps in this direction. Examples include the ranking of researchers using PageRank [29] or the beta distribution [30], and the re-scaling of citation distributions from different disciplines under a universal curve using the lognormal distribution [31].

One crucial aspect of the process of citation accumulation is that it takes a long time to reach a steady state [32]. This reality is often ignored in many analyses and thus confounds the interpretation of most measured values. Indeed, the lag between time of publication and perception of impact is becoming increasingly relevant. For example, faced with increasingly large pools of applicants, hiring committees need to be able to find the most qualified researchers for the position in an efficient and timely manner [33, 34]. To our knowledge, only a few attempts have been made in developing indicators that can predict future impact using citation measures [35, 36] and those have had limited success [37].

Here, we depart from previous efforts by developing a principled approach to the quantification of scientific impact. Specifically, we demonstrate that the distribution of the asymptotic number of accumulated citations to publications by a researcher or from a research institution is consistent with a discrete lognormal model [32, 38]. We validate our approach with two datasets acquired from Thomson Reuters’ Web of Science (WoS):

- Manually disambiguated citation data pertaining to researchers at the top United States (U.S.) research institutions across seven disciplines [39]: chemical engineering, chemistry, ecology, industrial engineering, material science, molecular biology, and psychology;
- Citation data from the chemistry departments of 106 U.S. institutions classified as “very high research activity”.

Significantly, our findings enable us to develop a measure of scientific impact with desirable properties.

## The Data

We perform our first set of analyses on the dataset described by Duch et al. [39]. This dataset contains the disambiguated publication records of 4,204 faculty members at some of the top U.S. research universities in seven scientific disciplines: chemical engineering, chemistry, ecology, industrial engineering, material science, molecular biology, and psychology (see [39] for details about data acquisition and validation). We consider here only 230,964 publications that were in press by the end of 2000. We do this so that every publication considered has had a time span of at least 10 years for accruing citations [38] (the researcher’s publication dataset was gathered in 2010).

We perform our second set of analyses on the publication records of the chemistry departments at the top U.S. research institutions according to [40]. Using the publications’ address fields, we identified 382,935 total publications from 106 chemistry departments that were in press by the end of 2009 (the department’s publication dataset was gathered in 2014).

In our analyses we distinguish between “primary” publications, which report original research findings, and “secondary” publications, which analyze, promote or compile research published elsewhere. We identify

as primary publications those classified by WoS as “Article”, “Letter”, or “Note” and identify all other publications types as secondary publications.

Moreover, to ensure that we have enough statistical power to determine the significance of the model fits, we restrict our analysis to researchers with at least 50 primary research publications. These restrictions reduce the size of the researchers dataset to 1,283 researchers and 148,878 publications. All 106 departments in our dataset have a total of more than 50 primary research publications.

## The Distribution of the asymptotic Number of Citations

Prior research suggests that a lognormal distribution can be used to approximate the steady-state citation profile of a researcher’s aggregated publications [31, 41]. Stringer et al. demonstrated that the distribution of the number  $n(t)$  of citations to publications published in a given journal in a given year converges to a stationary functional form after about ten years [32]. This result was interpreted as an indication that the publications published in a single journal have a characteristic citation propensity [42] which is captured by the distribution of the “ultimate” number of citations. Here, we investigate the asymptotic number of citations  $n_a$  to the publications of an individual researcher as well as the set of all researchers in a department at a research institution.

We hypothesize that  $n_a$  is a function of a latent variable  $\psi$  representing a publication’s “citability” [43]. The citability  $\psi$  results from the interplay of several, possibly independent, variables such as timeliness of the work, originality of approach, strength of conclusion, reputation of authors and journals, and potential for generalization to other disciplines, just to name a few [44, 45]. In the simplest case, citability will be additive in all these variables, in which case the applicability of the central limit theorem implies that  $\psi$  will be a Gaussian variable,  $\psi \in N(\mu_a, \sigma_a)$ , where  $\mu_a$  and  $\sigma_a$  are respectively the mean and standard deviation of the citability of the publications by researcher  $a$ . Therefore, the impact of a researcher’s body of work is described by a distribution characterized by just two parameters,  $\mu$  and  $\sigma$ . Similarly, because in the U.S. departments hire faculty based on their estimated quality, the researchers associated with a department will presumably be similar in stature or potential.

Unlike citations, which are observable and quantifiable, the variables contributing to  $\psi$  are neither easily observable nor easy to quantify. Moreover, mapping  $\psi$  into citations is not a trivial matter. Citation counts span many orders of magnitude, with the most highly cited publications having tens of thousands of citations [46]. Large-scale experiments on cultural markets indicate that social interactions often create a “rich get richer” dynamics, far distancing the quality of an underlying item from its impact [47]. Citation dynamics are no different. For example, Duch et al. recently showed that the  $h$ -index has a power-law dependence on the number of publications  $N_p$  of a researcher [39]. Here, we reduce the potential distortion of citation-accruing dynamics by focusing on the logarithm of  $n_a$ . In effect, we take  $n_a$  to be the result of a multiplicative process of the same variables determining  $\psi$ . Thus, we can calculate the probability  $p_{dln}(n_a)$  that a researcher or department will have a primary research publication with  $n_a$  citations, as an integral over  $\psi$ :

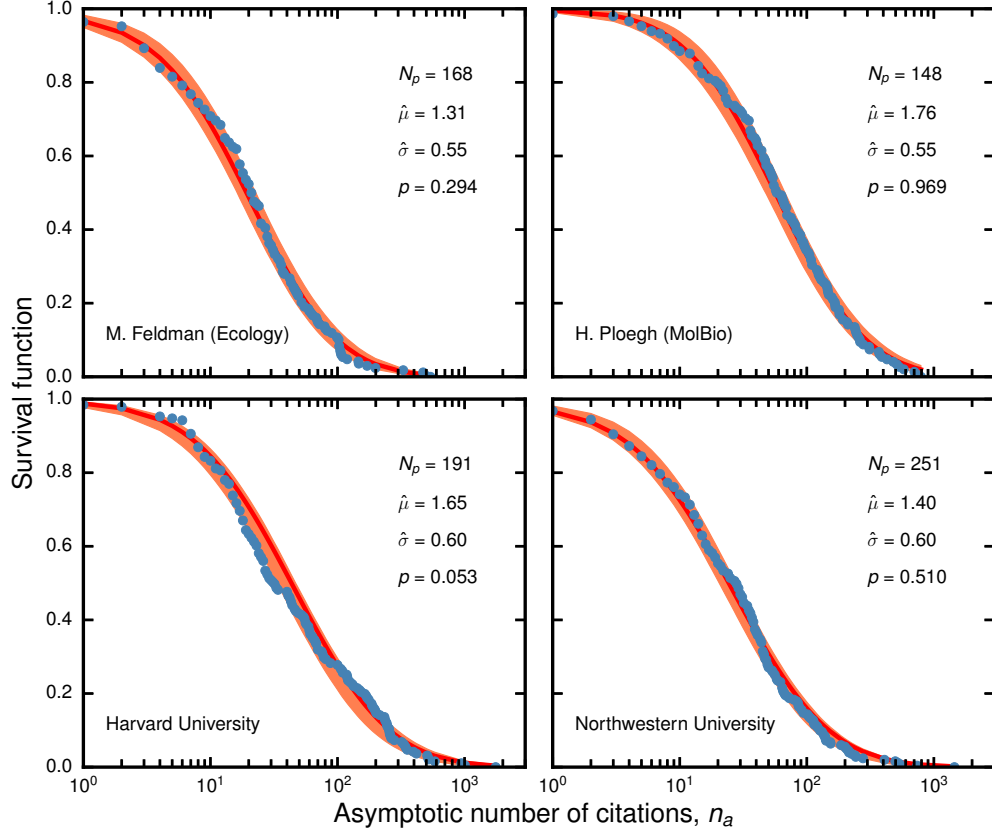
$$p_{dln}(n_a|\mu, \sigma) = \int_{\log_{10}(n_a)}^{\log_{10}(n_a+1)} \frac{d\psi}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\psi - \mu)^2}{2\sigma^2}\right) . \quad (1)$$

Most researchers also communicate their ideas to their peers via secondary publications such as conference proceedings which, in many disciplines, are mainly intended to promote related work published elsewhere. Some secondary publications will have significant timeliness, in particular review papers and editorial materials, and therefore will likely be cited too. Most of them, however, will not be cited at all. If accounting for secondary publications, Eq. (1) has to be generalized as:

$$P(n_a|\mu, \sigma, f_s, \boldsymbol{\theta}) = (1 - f_s)p_{dln}(n_a|\mu, \sigma) + f_s p_s(n_a|\boldsymbol{\theta}) , \quad (2)$$

where  $f_s$  is the fraction of secondary publications in a body of work and  $p_s(n_a|\boldsymbol{\theta})$  represents the probability distribution, characterized by parameters  $\boldsymbol{\theta}$  and not necessarily lognormal, of  $n_a$  for secondary research publications. We found that in practice Eq. (2) can be well approximated by:

$$P(n_a|\mu', \sigma', f_s) = f_s \delta_{0,n_a} + (1 - f_s)p_{dln}(n_a|\mu', \sigma') ,$$



**Figure 1. Distribution of the asymptotic number of citations to publications for researchers and chemistry departments in our database.** We fit Eq. (1) to all citations accrued by 2010 to publications published by 2000 for two researchers (**top row**), and to all citations accrued by 2013 to publications published in 2000 for two chemistry departments (**bottom row**). The red line shows the maximum likelihood fit of Eq. (1) to the data (blue circles). The light red region represents the 95% confidence interval estimated using bootstrap (1000 generated samples per empirical data point). We also show the number of publications  $N_p$  in each set and the parameter values of the individual fits.

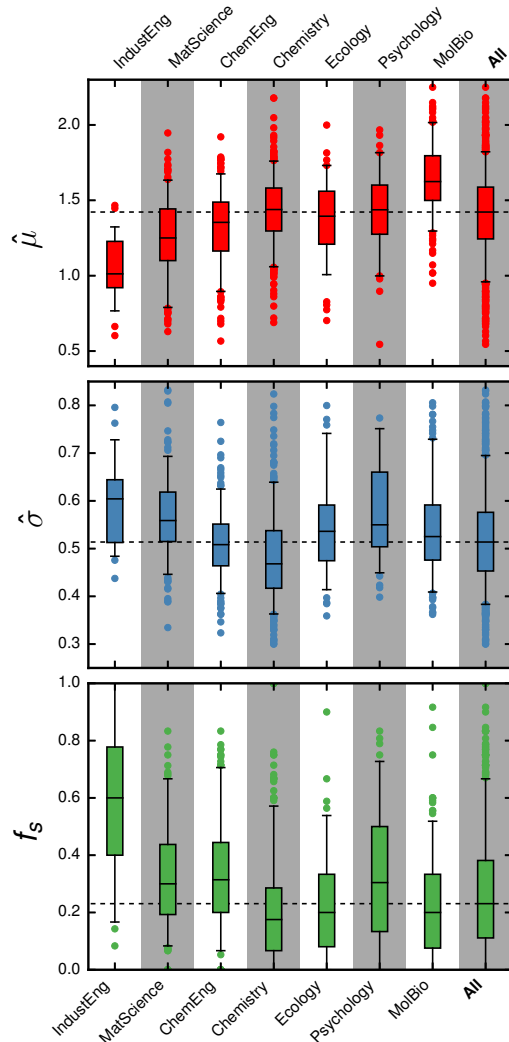
where  $\delta$  is the Kronecker delta. Surprisingly, we found that  $\mu' \approx \mu$  and  $\sigma' \approx \sigma$ , suggesting that secondary publications have citation characteristics that are significantly different from those of primary publications.

## Results

Figure 1 shows the cumulative distribution of citations to primary research publications of two researchers in our database and two chemistry departments. Using a  $\chi^2$  goodness-of-fit test with re-sampling [48], we find that we can reject the discrete lognormal model, Eq. (1), for only 2.88% of researchers and 1.13% of departments in our database. The results of our statistical analysis demonstrate that a discrete lognormal distribution with parameters  $\mu$  and  $\sigma$  provides an accurate description of the distribution of the asymptotic number of citations for a researcher's body of work and for the publications from an academic department.

Figure 2 displays the sample characteristics of the fitted parameters. The median value of  $\hat{\mu}$  obtained for the different disciplines lies between 1.0 and 1.6. Using data reported in [28] we find a significant correlation ( $\tau_{Kendall} = 0.62$ ,  $p = 0.069$ ) between the median value of  $\hat{\mu}$  for a discipline and the total number of citation to journals in that discipline (Fig. 3). This correlation suggests that  $\hat{\mu}$  depends on the typical number of citations to publications within a discipline. This dependence on discipline size can in principle be corrected by a normalization factor [14, 31, 49].

We also plot the fraction of secondary publications,  $f_s$ , for all the researchers. We find that nearly a



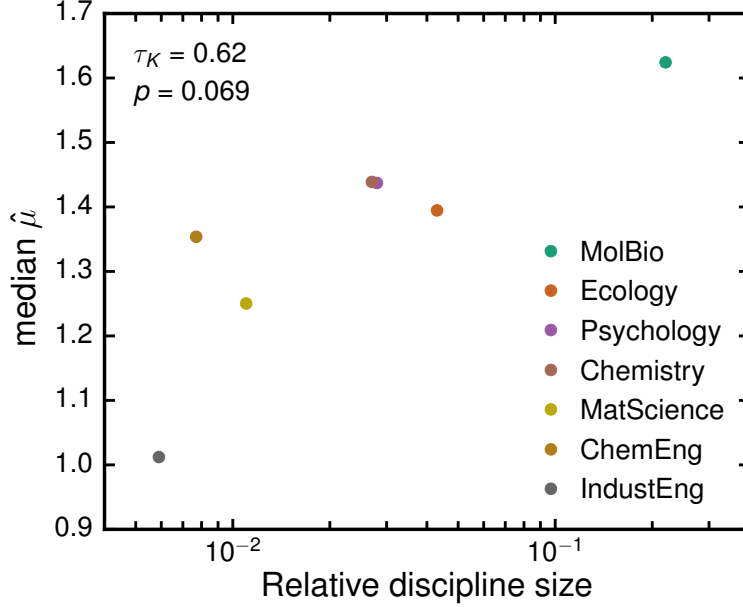
**Figure 2. Parameter statistics of all 1,283 researchers in the database grouped by discipline.** We show the maximum likelihood fitted model parameters (**top** and **center**) and the fraction of secondary publications (**bottom**). The black horizontal dashed line indicates the median of all researchers. For clarity, we do not show the values of  $\hat{\sigma}$  for 9 researchers that are outliers.

fourth of the publications of half of all researchers are secondary, but intra-discipline variation is high. Inter-discipline variability is also high: 17% of the publications of a typical researcher in chemistry are secondary, whereas 60% of the publications of a typical researcher in industrial engineering are secondary.

## Reliability of Estimation

We next investigate the dependence of the parameter estimates on number of publications,  $N_p$ , both at the individual level – testing the effect of sample size – and at the discipline level – testing overall dependence on  $N_p$ . To test for sample size dependence, we fit the model to subsets of a researcher’s publication list. We find that estimates of  $\sigma$  are more sensitive to sample size than estimates of  $\mu$  (Figs. S1 and S2). However, this dependence becomes rapidly negligible as the sample size approaches the minimum number of publications we required in creating our sample ( $N_p \geq 50$ ).

Next, we test whether, at the discipline level, there is any dependence of  $\hat{\mu}$  on  $N_p$ . We find no statistically significant correlation, except for a very weak dependence ( $R^2 \sim 0.035$ ,  $p = 0.0052$ ) of  $\hat{\sigma}$  on  $N_p$  for chemical engineering (Table S1). This is in stark contrast with the  $h$ -index which exhibits a marked dependence on



**Figure 3. Correlation between median  $\hat{\mu}$  for a discipline and the discipline’s relative size.** We use Rosvall et al. [28] reported values of the relative number of citations to publications in journals of several disciplines as a proxy for relative field size and compare them with the median value of  $\hat{\mu}$  in each discipline. A Kendall rank-correlation test yields a  $\tau_K = 0.62$  with  $p = 0.069$ . This correlation suggests that  $\hat{\mu}$  depends on the typical number of citations of a discipline.

number of publications [16].

Then, we test for variation of the estimated parameter values along a researcher’s career. To this end, we order each researcher’s publication records chronologically and divide them into three sets with equal number of publications and fitted the model to each set of publications. Each set represents the citability of the publications authored at a particular career stage of a researcher. Time trends in the estimated values of  $\mu$  would indicate that the citability of a researcher’s work changes over time. We find such a change for 25% of all researchers. For over 64% of those researchers whose citability changes of over time we find that  $\hat{\mu}$  increases (Table 1).

In general, a department has many more publications than any single researcher. Thus, we are able to apply the model from Eq. (1) to each year’s worth of departmental publications. This fine temporal resolution enables us to investigate whether there is any time-dependence in the citability of the publications from a department. Figure 4 shows the time-evolution of  $\hat{\mu}$  for the chemistry departments at four typical research institutions. We see that both  $\hat{\mu}$  (circles) and  $\hat{\sigma}$  (vertical bars) remain remarkably stable over the period considered.

## Development of an Indicator

In the following, we compare the effectiveness of  $\mu$  as an impact indicator with that of other indicators. First, we test the extent to which the value of  $\mu_i$  for a given researcher is correlated with the values of other indicators for the same researcher. In order to provide an understanding of how the number of publications  $N_p$  influences the values of other metrics, we generate thousands of synthetic samples of  $n_a$  for different values of  $N_p$  and  $\mu_i$ , and a fixed value of  $\sigma$  for each discipline. We find that  $\mu$  is tightly correlated with several other measures, especially with the median number of citations (Fig. 5). Indeed  $\hat{\mu}$  can be estimated from the median number of citations:

$$\hat{\mu} \cong \log_{10}[\text{median}(n_a)] \quad , \quad (3)$$

This close relation between mean and logarithm of the median further supports our hypothesis of a lognormal distribution for the asymptotic number of citations to primary publications by a researcher.

Discipline	Upward trend in $\hat{\mu}$	Downward trend in $\hat{\mu}$
ChemEng	12%	19%
Chemistry	26%	6%
Ecology	8%	8%
IndustEng	0%	33%
MatScience	10%	11%
MolBio	5%	8%
Psychology	0%	0%
<b>All</b>	<b>16%</b>	<b>9%</b>

**Table 1. Trends of  $\hat{\mu}$  on career stage for the seven disciplines considered.** We divide each researcher’s chronologically-ordered publication records into three sets with equal number of publications (start, middle, and end) and fit the model to each set of publications to obtain  $\hat{\mu}^s$ ,  $\hat{\mu}^m$ , and  $\hat{\mu}^e$ . We then used ordinary-least-squares to perform a linear regression on the time dependence of  $(\hat{\mu}^s, \hat{\mu}^m, \hat{\mu}^e)$ . We then calculate the fraction of researchers whose  $\mu$  exhibits a statistically significant dependence on career length, by performing a two-tailed significance test on the slope of the regression. We use a randomization test (1,000 samples), combined with a multiple hypothesis correction [50] (*false discovery rate* of 0.05) to calculate a  $p$ -value: for each researcher, we randomly re-order his or her publications, divide them into three sets with equal number of publications and fit the model to each set of publications, and calculate the new slope; we obtain a  $p$ -value by comparing the original slope of the fit with the distribution of the randomized slopes.

An important factor to consider when designing a bibliometric indicator is its susceptibility to manipulation. Both the number of publications and total or average number of citations are easily manipulated, especially with the ongoing proliferation of journals of dubious reputation [51, 52]. Indeed, the  $h$ -index was introduced as a metric that resists manipulation. However, it is a straightforward exercise to show that one could achieve  $h \propto \sqrt{N_p}$  exclusively through self-citations. Indeed, because the  $h$ -index does not account for the effect of self-citations, it is rather susceptible to manipulation, especially by researchers with low values of  $h$  [53, 54].

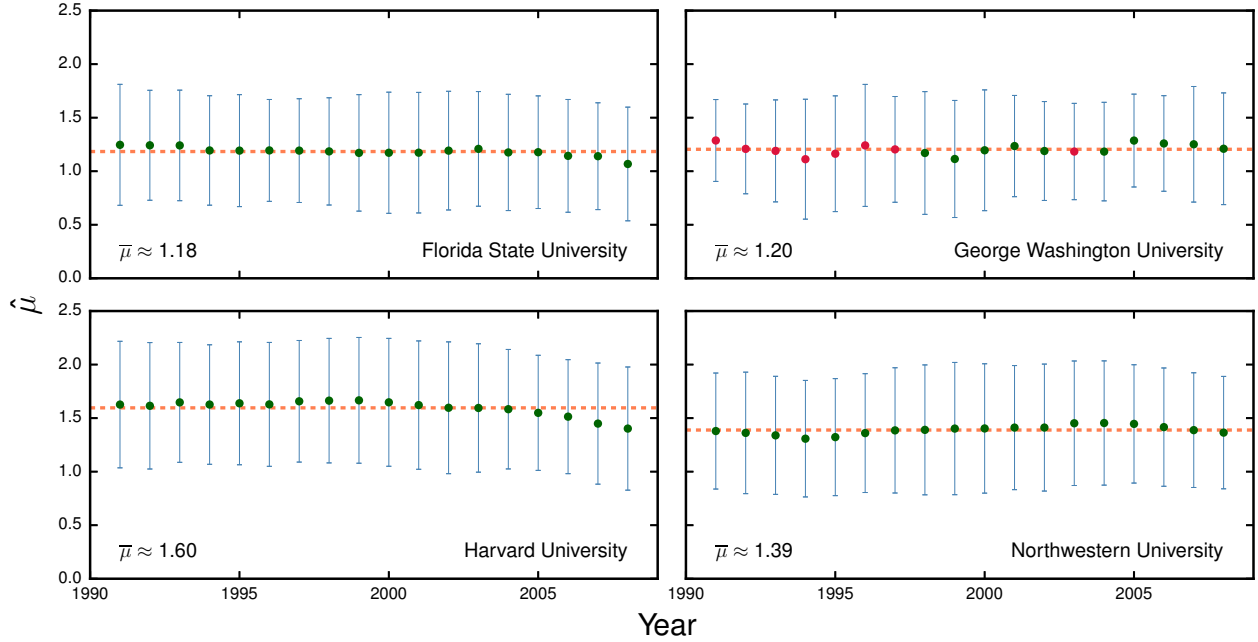
In order to determine the true susceptibility of the  $h$ -index to manipulation, we devise a method to raise a researcher’s  $h$ -index using the least possible number of self-citations (see Materials and Methods for details). Our results suggest that increasing the  $h$ -index by a small amount is no hard feat for researchers with the ability to quickly produce new articles (Fig. 6A).

Our proposed indicator,  $\mu$ , is far more difficult to manipulate. Because it has a more complex dependence on the number of citations than the  $h$ -index, to increase  $\mu$  in an efficient manner we use a process whereby we attempt to increase the median number of citations of a researcher’s work (see Materials and Methods for details). Specifically, we manipulated  $\mu$  for all the researchers by increasing their median number of citations. Remarkably, to increase  $\mu$  by a certain factor one needs at least 10 times more self-citations than one would need in order to increase the  $h$ -index by the same factor (Fig. 6B).

While a difference of 2 to 3 orders of magnitude in number of required self-citations may seem surprising for a measure so correlated with citation numbers (Fig. 5), the fact that  $\hat{\mu}$  is actually dependent on the citations to half of all primary publications by a researcher (Eq. (3)) makes  $\hat{\mu}$  less susceptible than the  $h$ -index to manipulation of citation counts from a small number of publications. This view is also supported by the fact that increasing citations may actually decrease  $\hat{\mu}$ , as we may be adding them to a publication that would not be expected to receive that number of citations given the lognormal model. As a result, manipulation of scientific performance would be very difficult if using a  $\mu$ -based index.

## Comparison of Parameter Statistics

Finally we estimate the parameters in Eq. (1) for chemistry journals and compare  $\hat{\mu}$  of chemistry departments and journals in selected years, and all chemistry researchers in our database (Fig. 7. See Fig. S4 for  $\hat{\sigma}$  and  $f_s$  comparison). In order to make sense of this comparison, we must note a few aspects about the data. The researchers in the database were affiliated with the top 30 chemistry departments in the U.S., whereas the set of chemistry departments covers all the chemistry departments from very high research activity universities.



**Figure 4. Time-evolution of departments  $\hat{\mu}$ .** Each circle and bar represent, respectively, the  $\hat{\mu}$  and  $\hat{\sigma}$  for a given year of publications. We estimate the parameters in Eq. (1) for sets of departmental publications using a “sliding window” of 3 years. Fits for which we cannot reject the hypothesis that the data is consistent with a discrete lognormal distribution are colored green. We also show each department’s average value of  $\hat{\mu}$  over the period considered (orange dashed lines).

Thus, it is natural that the typical  $\hat{\mu}$  of researchers is higher than that of departments. Not surprisingly, we find that  $\hat{\mu}$  is typically the lowest for journals.

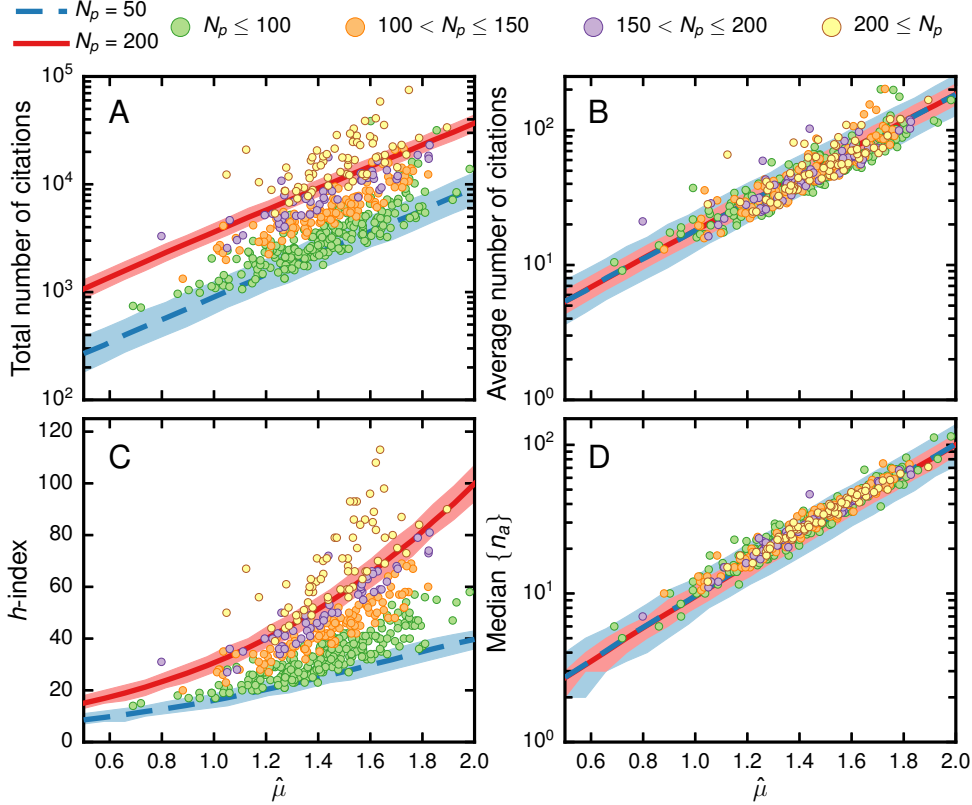
## Discussion

The ever-growing size of the scientific literature precludes researchers from following all developments from even a single sub-field. Therefore researchers need proxies of quality in order to identify which publications to browse, read, and cite. Three main heuristics are familiar to most researchers: institutional reputation, journal reputation, and author reputation.

Author reputation has the greatest limitations. Researchers are not likely to be known outside their (sub-)field and young researchers will not even be known outside their labs. Similarly, if we exclude a few journals with multidisciplinary reputations (Nature, Science, PNAS, NEJM), the reputation of a scientific journal is unlikely to extend outside its field. Institutional reputations are the most likely to be known broadly. Cambridge, Harvard, Oxford, and Stanford are widely recognized. However, one could argue that institutional reputation is not a particularly useful heuristic for finding quality publications within a specific research field.

Our results show that the expected citability of scientific publications published by (i) the researchers in a department, (ii) a given scientific journal, or (iii) a single researcher can be set on the single scale defined by  $\mu$ . Thus, for a researcher whose publications are characterized by a very high  $\mu$ , authorship of a publication may give a stronger quality signal about the publication than the journal in which the study is being published. Conversely, for an unknown researcher the strongest quality signal is likely to be the journal where the research is being published or the institution the researcher is affiliated with. Our results thus provide strong evidence for the validity of the heuristics used by most researchers and clarify the conditions under which they are appropriate.



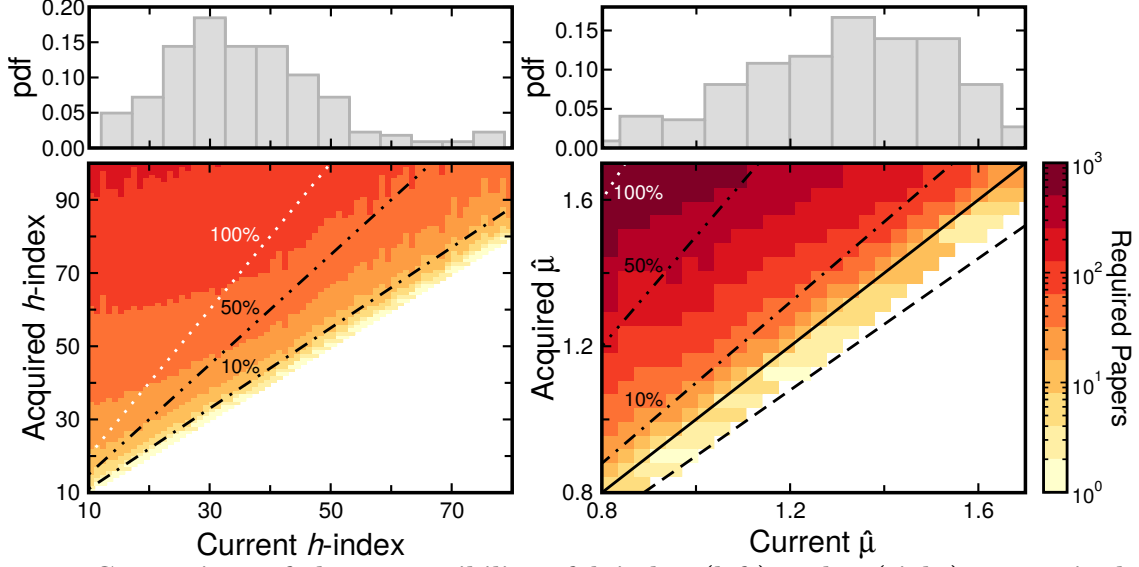


**Figure 5. Dependence of popular impact metrics on the values of  $\hat{\mu}$  and number of publications  $N_p$  for researchers in chemistry.** We generate 1000 synthetic datasets for each of 20 values of  $\hat{\mu}$  from 0.5 to 2.0, inclusive, and for  $N_p = 50$  (blue) and  $N_p = 200$  (red). We use the average  $\hat{\sigma}$  of all researchers in chemistry. For each pair of values of  $\hat{\mu}$  and  $N_p$  we calculated the average value and 95% confidence interval. The colored circles indicate the observed values of the corresponding metrics for chemistry, which have been grouped according to their number of publications  $N_p$ . Values for 22 researchers fall outside of the figures' limits: 3 in A, 7 in B, 4 in C, 3 in D. (A) The total number of citations depends dramatically on  $N_p$ , which in turn depends strongly on career length, and can be influenced by just a few highly cited publications. (B) The average number of citations is less susceptible to changes in  $N_p$  but can still be influenced by a small number of highly cited publications. (C) The  $h$ -index, like the total number of publications, is strongly dependent on  $N_p$ . (D) The median number of citations to publications, like the average, is not very dependent on  $N_p$ , and can capture most of the observed behavior.

## Materials and Methods

### Model Fitting and Hypothesis Testing

We estimate the discrete lognormal model parameters of Eq. (1) for all 1,283 researchers in our database using a maximum likelihood estimator [38]. We then test the goodness of the fit, at an individual level using the  $\chi^2$  statistical test. We bin the empirical data in such a way that there are at least 5 expected observation per bin. To assess significance we calculate the  $\chi_o^2$  statistic for each researcher and then, for each of them, re-sample their citation records using bootstrap (1,000 samples) and calculate a new value of the statistics  $\chi_i^2$  ( $i = 1, \dots, 1,000$ ). We then extract a p-value by comparing the observed statistic  $\chi_o^2$  with the re-sampled  $\chi^2$  distribution. Finally we use a multiple hypothesis correction [50], with a *false discovery rate* of 0.05, when comparing the model fits with the null hypothesis.



**Figure 6. Comparison of the susceptibility of  $h$ -index (left) and  $\mu$  (right) to manipulation.** **Bottom panel:** For each researcher in the database, we add publications with self-citations until we reach the desired value of index (see main text for details). The dashed black, dotted-dashed black and dotted white lines indicate the number of publications required to increase the index value by 10%, 50% and 100%, respectively. The solid diagonal black line indicates when the current value of  $\hat{\mu}$  is equal to the manipulated  $\hat{\mu}$ . The dark blue vertical line represents the average value of the indicator amongst all researchers in our database. **Top panel:** Distributions of current  $h$ -index (left) and  $\hat{\mu}$  (right) for all researchers in the database.

## Generation of Theoretical Performance Indicators

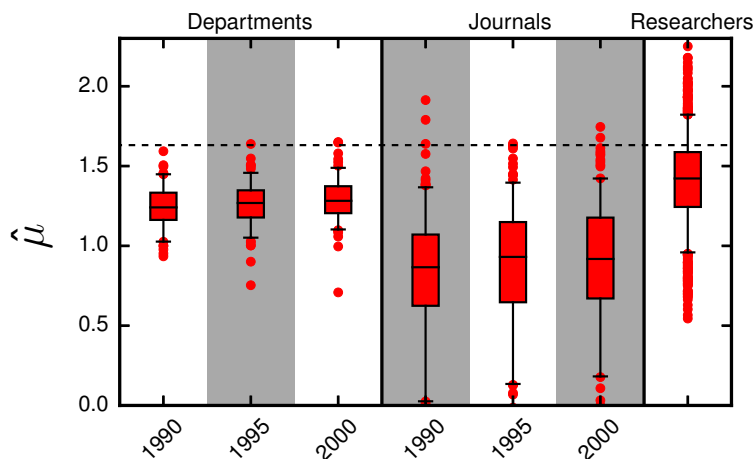
For each discipline we take the average value of  $\hat{\sigma}$  and 20 equally spaced values of  $\mu$  between 0.5 and 2.0. We then generate 1,000 datasets of 50 and 200 publications by random sampling from Eq. (1). We then fit the model individually to these 2,000 synthetic datasets and extracted the  $h$ -index, average number of citations, total number of citations and median number of citations to publications with at least one citation. Finally, for each value of  $\mu$ , we calculate the average and the 95% confidence interval of all the indicators.

## Manipulation Procedure for $h$ -index

We try to increase the  $h$ -index of a researcher by self-citations alone, i.e., we assume the researcher does not receive citations from other sources during this procedure. The procedure works by adding only the minimum required citations to those publications that would cause the  $h$ -index to increase. Consider researcher John Doe who has 3 publications with  $\{n_a\} = (2, 2, 5)$ . Doe's  $h$  is 2. Assuming those publications don't get cited by other researchers during this time period, to increase  $h$  by 1, Doe needs to publish only one additional publication with two self-citations; to increase  $h$  by 2 he must instead produce five publications with a total of eight self-citations, four of which to one of the additional five publications. We execute this procedure for all researchers in the database until they reached a  $h$ -index of 100.

## Manipulation Procedure for $\mu$

The manipulation of  $\mu$  is based on Eq. (3). We try to change a researcher's  $\mu$  by increasing the median number of citations to publications which have at least one citation already. We consider only self-citations originating from secondary publications, i.e., publications that will not get cited. For a given corpus of publications we first define a target increase in median,  $x$  and then calculate the number of self-citations needed to increase the current median by  $x$  citations and the corresponding number of secondary publications. We then take the initial corpus of publications and attempt to increase the median citation by  $x + 1$ . We repeat this procedure until we reach an increase in median citation of 2000.



**Figure 7. Comparison of  $\hat{\mu}$  across departments, journals, and researchers.** We show the maximum likelihood fitted  $\hat{\mu}$  for chemistry departments and chemistry journals in select years, and for all chemistry researchers in our database. The black horizontal dashed lines mark the value of the corresponding parameter for the *Journal of the American Chemical Society* in 1995. For clarity, we do not show  $\hat{\mu}$  for 23 journals that are outliers.

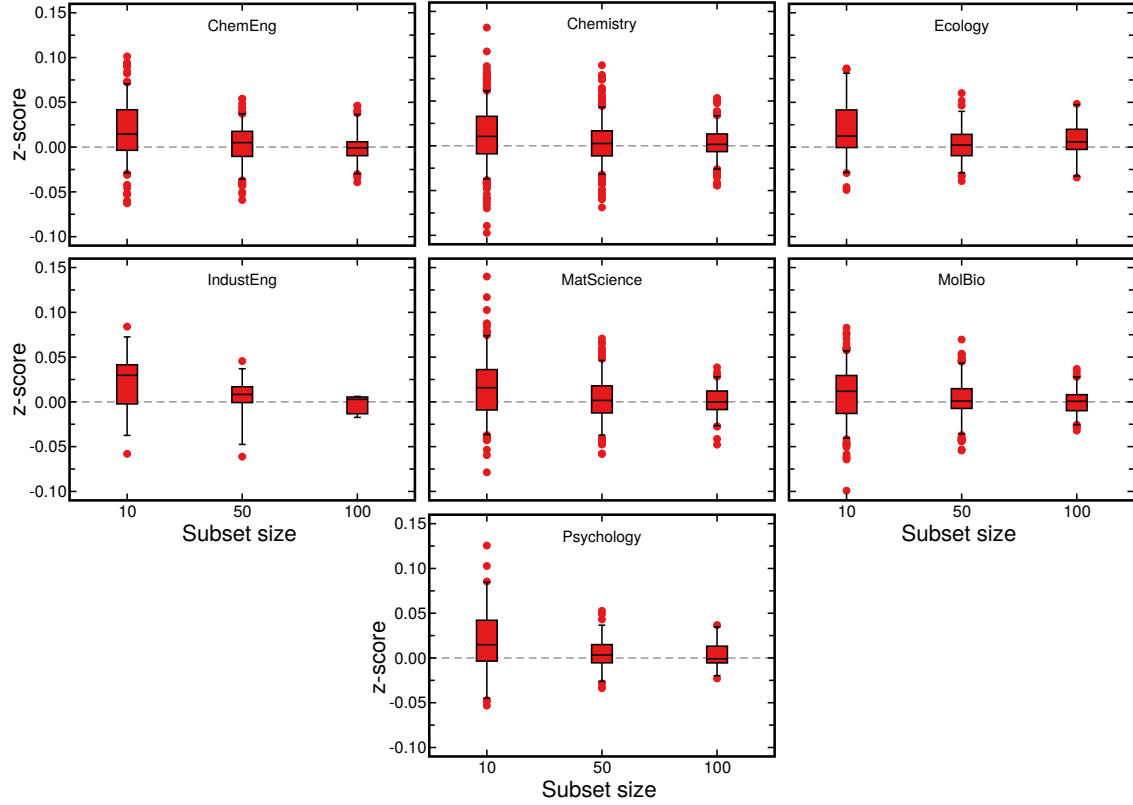
## References

- [1] David Nicholas, Anthony Watkinson, Rachel Volentine, Suzie Allard, Kenneth Levine, Carol Tenopir, and Eti Herman. Trust and Authority in Scholarly Communications in the Light of the Digital Transition: setting the scene for a major study. *Learn. Publ.*, 27(2):121–134, April 2014. ISSN 09531513. doi: 10.1087/20140206.
- [2] Eugene Garfield and Irving H. Sher. *Genetics Citation Index*. Institute for Scientific Information, Philadelphia, 1963.
- [3] Eugene Garfield. The History and Meaning of the Journal Impact Factor. *JAMA J. Am. Med. Assoc.*, 295(1):90, January 2006. ISSN 0098-7484. doi: 10.1001/jama.295.1.90.
- [4] Jorge E. Hirsch. An index to quantify an individual’s scientific research output. *Proc. Natl. Acad. Sci.*, 102(46):16569–16572, November 2005. ISSN 0027-8424. doi: 10.1073/pnas.0507655102.
- [5] Michael H. MacRoberts and Barbara R MacRoberts. Problems of citation analysis: A critical review. *J. Am. Soc. Inf. Sci.*, 40(5):342–349, September 1989. ISSN 0002-8231. doi: 10.1002/(SICI)1097-4571(198909)40:5<342::AID-ASI7>3.0.CO;2-U.
- [6] F. Narin and Kimberly S. Hamilton. Bibliometric performance measures. *Scientometrics*, 36(3):293–310, July 1996. ISSN 0138-9130. doi: 10.1007/BF02129596.
- [7] Jonathan R. Cole. A Short History of the Use of Citations as a Measure of the Impact of Scientific and Scholarly Work. In *web Knowl. A Festschrift Honor Eugene Garf.*, chapter 14, pages 281–300. Information Today, 2000.
- [8] Wolfgang Glänzel and Henk F Moed. Journal impact measures in bibliometric research. *Scientometrics*, 53(2):171–193, 2002. ISSN 01389130. doi: 10.1023/A:1014848323806.
- [9] Christine L. Borgman and Jonathan Furner. Scholarly communication and bibliometrics. *Annu. Rev. Inf. Sci. Technol.*, 36(1):2–72, February 2005. ISSN 00664200. doi: 10.1002/aris.1440360102.
- [10] Peter Vinkler. Characterization of the impact of sets of scientific papers: The Garfield (impact) factor. *J. Am. Soc. Inf. Sci. Technol.*, 55(5):431–435, March 2004. ISSN 1532-2882. doi: 10.1002/asi.10391.

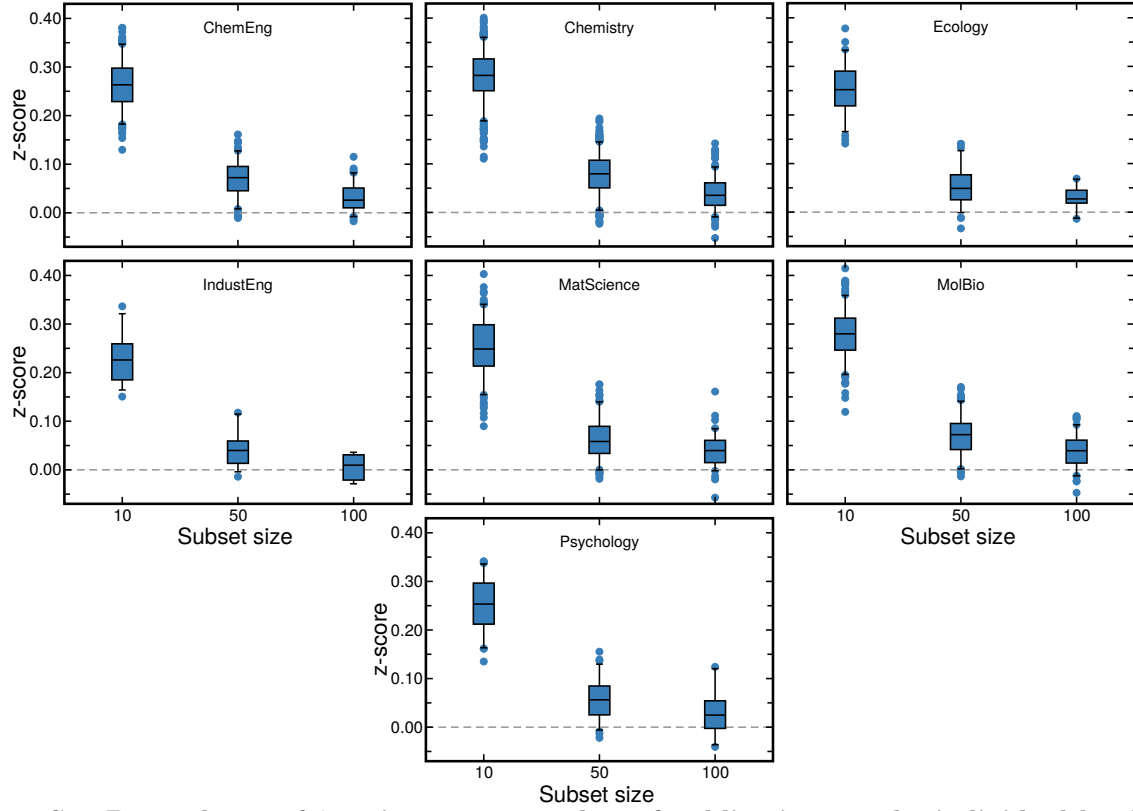
- [11] Lutz Bornmann and Hans-Dieter Daniel. What do we know about the index? *J. Am. Soc. Inf. Sci. Technol.*, 58(9):1381–1385, July 2007. ISSN 15322882. doi: 10.1002/asi.20609.
- [12] Lutz Bornmann and Hans-Dieter Daniel. What do citation counts measure? A review of studies on citing behavior. *J. Doc.*, 64(1):45–80, January 2008. ISSN 0022-0418. doi: 10.1108/00220410810844150.
- [13] S. Alonso, F.J. Cabrerizo, E. Herrera-Viedma, and F. Herrera. h-Index: A review focused in its variants, computation and standardization for different scientific fields. *J. Informetr.*, 3(4):273–289, October 2009. ISSN 17511577. doi: 10.1016/j.joi.2009.04.001.
- [14] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. Statistical physics of social dynamics. *Rev. Mod. Phys.*, 81(2):591–646, May 2009. ISSN 0034-6861. doi: 10.1103/RevModPhys.81.591.
- [15] Allen W. Wilhite and Eric A. Fong. Coercive Citation in Academic Publishing. *Science*, 335(6068):542–543, February 2012. ISSN 0036-8075. doi: 10.1126/science.1212540.
- [16] Leo Egghe. Dynamich-index: The Hirsch index in function of time. *J. Am. Soc. Inf. Sci. Technol.*, 58(3):452–454, February 2007. ISSN 15322882. doi: 10.1002/asi.20473.
- [17] Leo Egghe. Theory and practise of the g-index. *Scientometrics*, 69(1):131–152, October 2006. ISSN 0138-9130. doi: 10.1007/s11192-006-0144-7.
- [18] BiHui Jin, LiMing Liang, Ronald Rousseau, and Leo Egghe. The R- and AR-indices: Complementing the h-index. *Chinese Sci. Bull.*, 52(6):855–863, March 2007. ISSN 1001-6538. doi: 10.1007/s11434-007-0145-9.
- [19] Fiorenzo Franceschini, Domenico Maisano, Anna Perotti, and Andrea Proto. Analysis of the ch-index: an indicator to evaluate the diffusion of scientific research output by citers. *Scientometrics*, 85(1):203–217, October 2010. ISSN 0138-9130. doi: 10.1007/s11192-010-0165-0.
- [20] Laura Bonetta. Should You Be Tweeting? *Cell*, 139(3):452–453, October 2009. ISSN 00928674. doi: 10.1016/j.cell.2009.10.017.
- [21] Sibeles Fausto, Fabio a Machado, Luiz Fernando J Bento, Atila Iamarino, Tatiana R Nahas, and David S Munger. Research Blogging: Indexing and Registering the Change in Science 2.0. *PLoS One*, 7(12):e50109, December 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0050109.
- [22] Roberta Kwok. Research impact: Altmetrics make their mark. *Nature*, 500(7463):491–493, August 2013. ISSN 0028-0836. doi: 10.1038/nj7463-491a.
- [23] Stefanie Haustein and Tobias Siebenlist. Applying social bookmarking data to evaluate journal usage. *J. Informetr.*, 5(3):446–457, May 2011. ISSN 17511577. doi: 10.1016/j.joi.2011.04.002.
- [24] Jason Priem, Heather A. Piwowar, and Bradley M. Hemminger. Altmetrics in the wild: Using social media to explore scholarly impact. *arXiv1203.4745v1 csDL 20 Mar 2012*, 1203.4745:1–23, March 2012. ISSN null. doi: <http://arxiv.org/abs/1203.4745v1>.
- [25] Amélie Yavchitz, Isabelle Boutron, Aida Bafeta, Ibrahim Marroun, Pierre Charles, Jean Mantz, and Philippe Ravaud. Misrepresentation of Randomized Controlled Trials in Press Releases and News Coverage: A Cohort Study. *PLoS Med.*, 9(9):e1001308, September 2012. ISSN 1549-1676. doi: 10.1371/journal.pmed.1001308.
- [26] Philip M. Davis and Jason S. Price. eJournal interface can influence usage statistics: Implications for libraries, publishers, and Project COUNTER. *J. Am. Soc. Inf. Sci. Technol.*, 57(9):1243–1248, July 2006. ISSN 15322882. doi: 10.1002/asi.20405.
- [27] Terrence A. Brooks. Evidence of complex citer motivations. *J. Am. Soc. Inf. Sci.*, 37(1):34–36, 1986. ISSN 0002-8231. doi: 10.1002/(sici)1097-4571(198601)37:1%3C34::aid-asi5%3E3.0.co;2-0.

- [28] Martin Rosvall and Carl T Bergstrom. Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci.*, 105(4):1118–1123, January 2008. ISSN 0027-8424. doi: 10.1073/pnas.0706851105.
- [29] Filippo Radicchi, Santo Fortunato, Benjamin Markines, and Alessandro Vespignani. Diffusion of scientific credits and the ranking of scientists. *Phys. Rev. E*, 80(5):056103, November 2009. ISSN 1539-3755. doi: 10.1103/PhysRevE.80.056103.
- [30] Alexander M. Petersen, H. Eugene Stanley, and Sauro Succi. Statistical regularities in the rank-citation profile of scientists. *Sci. Rep.*, 1:181, December 2011. ISSN 2045-2322. doi: 10.1038/srep00181.
- [31] Filippo Radicchi, Santo Fortunato, and Claudio Castellano. Universality of citation distributions: Toward an objective measure of scientific impact. *Proc. Natl. Acad. Sci.*, 105(45):17268–17272, November 2008. ISSN 0027-8424. doi: 10.1073/pnas.0806977105.
- [32] Michael J. Stringer, Marta Sales-Pardo, and Luís A. Nunes Amaral. Effectiveness of Journal Ranking Schemes as a Tool for Locating Information. *PLoS One*, 3(2):e1683, February 2008. ISSN 1932-6203. doi: 10.1371/journal.pone.0001683.
- [33] Sune Lehmann, Andrew D. Jackson, and Benny E. Lautrup. Measures for measures. *Nature*, 444(7122):1003–1004, December 2006. ISSN 0028-0836. doi: 10.1038/4441003a.
- [34] Alison Abbott, David Cyranoski, Nicola Jones, Brendan Maher, Quirin Schiermeier, and Richard Van Noorden. Metrics: Do metrics matter? *Nature*, 465(7300):860–862, June 2010. ISSN 0028-0836. doi: 10.1038/465860a.
- [35] Daniel E Acuna, Stefano Allesina, and Konrad P. Kording. Future impact: Predicting scientific success. *Nature*, 489(7415):201–202, September 2012. ISSN 0028-0836. doi: 10.1038/489201a.
- [36] Amin Mazloumian. Predicting Scholars’ Scientific Impact. *PLoS One*, 7(11):e49246, November 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0049246.
- [37] Orion Penner, Raj K Pan, Alexander M. Petersen, Kimmo Kaski, and Santo Fortunato. On the Predictability of Future Impact in Science. *Sci. Rep.*, 3:3052, October 2013. ISSN 2045-2322. doi: 10.1038/srep03052.
- [38] Michael J. Stringer, Marta Sales-Pardo, and Luís A. Nunes Amaral. Statistical validation of a global model for the distribution of the ultimate number of citations accrued by papers published in a scientific journal. *J. Am. Soc. Inf. Sci. Technol.*, 61(7):1377–1385, April 2010. ISSN 15322882. doi: 10.1002/asi.21335.
- [39] Jordi Duch, Xiao Han T Zeng, Marta Sales-Pardo, Filippo Radicchi, Shayna Otis, Teresa K Woodruff, and Luís A. Nunes Amaral. The Possible Role of Resource Requirements and Academic Career-Choice Risk on Gender Differences in Publication Rate and Impact. *PLoS One*, 7(12):e51332, December 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0051332.
- [40] List of research universities in the United States. URL [https://en.wikipedia.org/wiki/List\\_of\\_research\\_universities\\_in\\_the\\_United\\_States](https://en.wikipedia.org/wiki/List_of_research_universities_in_the_United_States).
- [41] Sidney Redner. Citation statistics from 110 years of physical review. *Phys. Today*, 58(6):49–54, June 2005. ISSN 0031-9228. doi: 10.1063/1.1996475.
- [42] Quentin L. Burrell. Predicting future citation behavior. *J. Am. Soc. Inf. Sci. Technol.*, 54(5):372–378, March 2003. ISSN 1532-2882. doi: 10.1002/asi.10207.
- [43] Quentin L. Burrell. Stochastic modelling of the first-citation distribution. *Scientometrics*, 52(1):3–12, 2001. ISSN 01389130. doi: 10.1023/A:1012751509975.
- [44] William Shockley. On the Statistics of Individual Variations of Productivity in Research Laboratories. *Proc. IRE*, 45(3):279–290, 1957. ISSN 0096-8390. doi: 10.1109/JRPROC.1957.278364.

- [45] Adrian Letchford, Helen Susannah Moat, and Tobias Preis. The advantage of short paper titles. *R. Soc. Open Sci.*, 2(8):150266, August 2015. ISSN 2054-5703. doi: 10.1098/rsos.150266.
- [46] Richard Van Noorden, Brendan Maher, and Regina Nuzzo. The top 100 papers. *Nature*, 514(7524): 550–553, October 2014. ISSN 0028-0836. doi: 10.1038/514550a.
- [47] Matthew J Salganik. Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market. *Science*, 311(5762):854–856, February 2006. ISSN 0036-8075. doi: 10.1126/science.1121066.
- [48] Bryan F.J. Manly. *Randomization, bootstrap and Monte Carlo methods in biology*. Chapman and Hall/CRC, 3rd edition, 2006. ISBN 1584885416.
- [49] Alexander M. Petersen, Fengzhong Wang, and H. Eugene Stanley. Methods for measuring the citations and productivity of scientists across time and discipline. *Phys. Rev. E*, 81(3):036114, March 2010. ISSN 1539-3755. doi: 10.1103/PhysRevE.81.036114.
- [50] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B*, 57(1):289–300, 1995.
- [51] John Bohannon. Who’s Afraid of Peer Review? *Science*, 342(6154):60–65, October 2013. ISSN 0036-8075. doi: 10.1126/science.342.6154.60.
- [52] Declan Butler. Investigating journals: The dark side of publishing. *Nature*, 495(7442):433–435, March 2013. ISSN 0028-0836. doi: 10.1038/495433a.
- [53] Michael Schreiber. Self-citation corrections for the Hirsch index. *Europhys. Lett.*, 78(3):30002, May 2007. ISSN 0295-5075. doi: 10.1209/0295-5075/78/30002.
- [54] Leif Engqvist and Joachim G. Frommen. The h-index and self-citations. *Trends Ecol. Evol.*, 23(5): 250–252, May 2008. ISSN 01695347. doi: 10.1016/j.tree.2008.01.009.

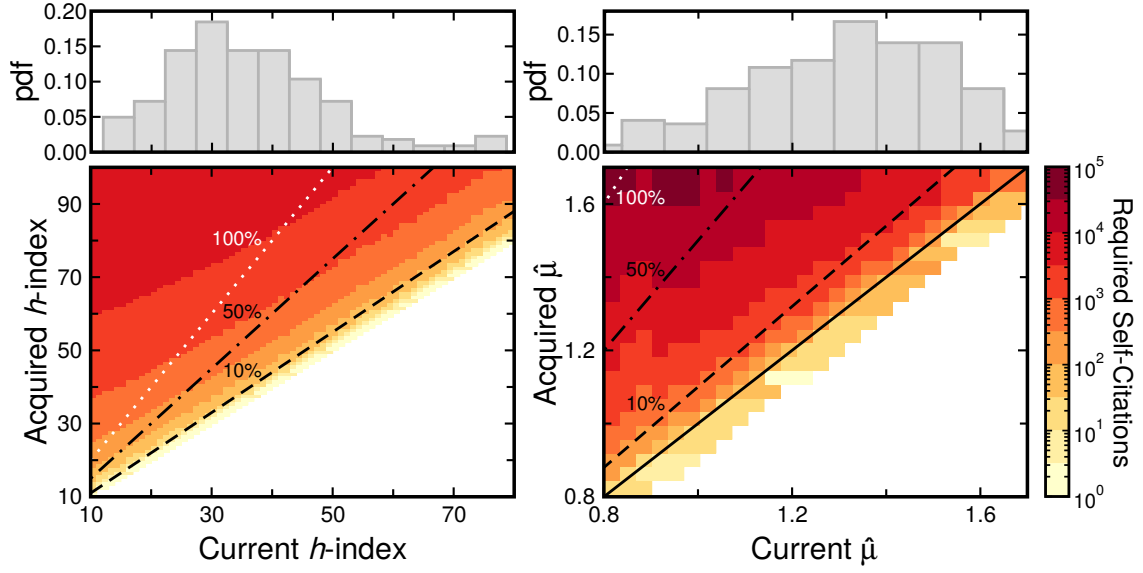


**Figure S1. Dependence of  $\hat{\mu}$  on number of publications at the individual level.** We fit the model to 1,000 randomized subsets of each researcher’s publication list and compare the  $\hat{\mu}$  obtained from fitting each subset of 10, 50, and 100 publications with the  $\hat{\mu}$  associated with the complete publication list. Then, for each researcher and subset size, we calculate a z-score using the mean and standard deviation of the “sub- $\hat{\mu}$ ”. For  $N_p \geq 50$ , the dependence on sample size is negligible for most researchers. Researchers with  $N_p < 100$  are omitted from the calculation on the subset of size 100.

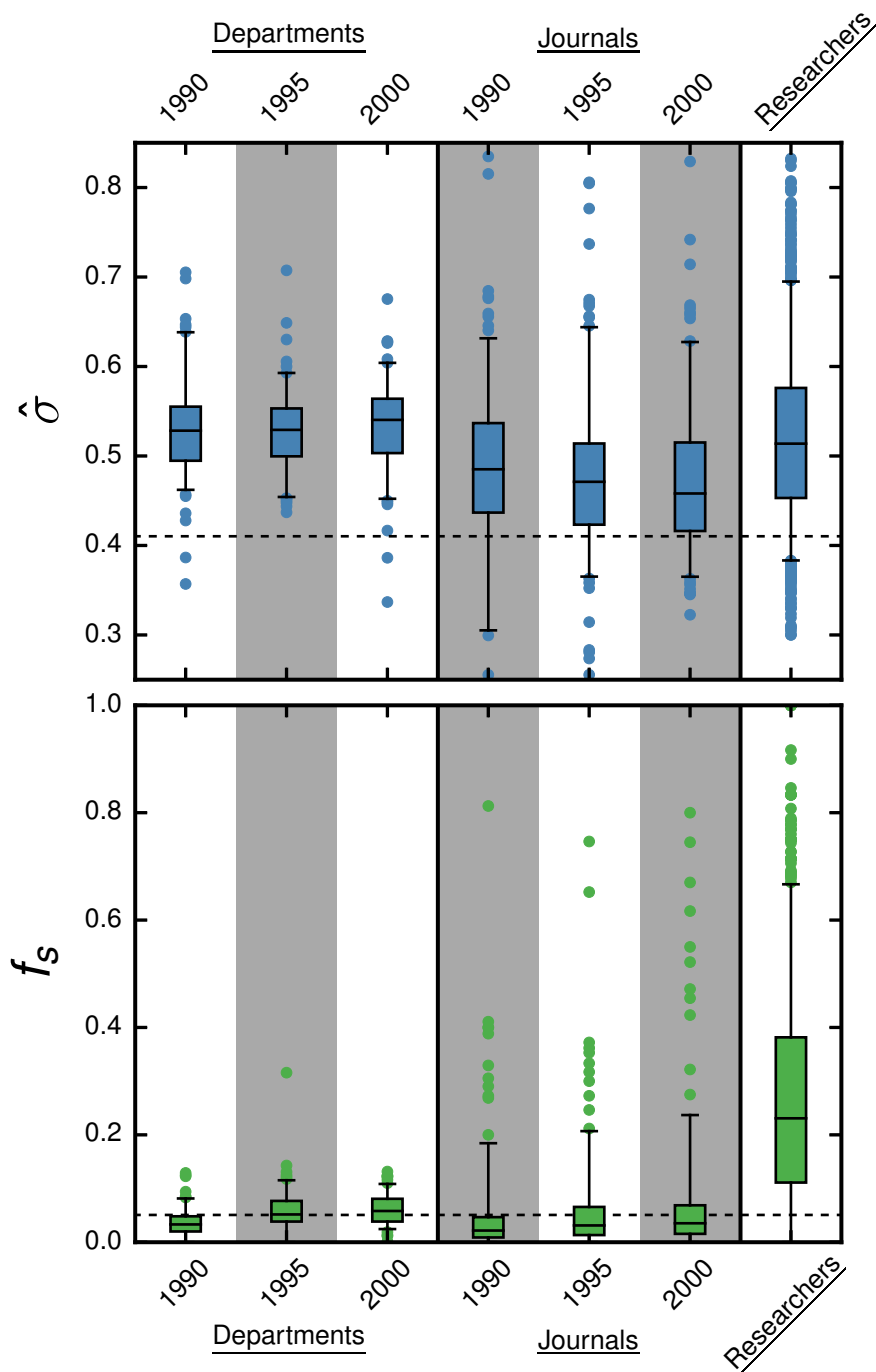


**Figure S2. Dependence of  $\hat{\sigma}$  estimates on number of publications at the individual level.** We use the same procedure as in Fig. S1, except here we show the results for the dependence of  $\hat{\sigma}$  on sample size. Estimates of  $\hat{\sigma}$  are more dependent of sample size than  $\hat{\mu}$ . However, as in the case of  $\hat{\mu}$ , the dependence of  $\hat{\sigma}$  on sample size decays rapidly with increasing sample size. Researchers with  $N_p < 100$  are omitted from the calculation on the subset of size 100.





**Figure S3. Susceptibility of impact measures to manipulation.** We used the same procedure as in Fig. 6, except here we show the required number of publications with self-citations that researchers need to publish in order to increase their indicators. Other details are the same as in Fig. 6.



**Figure S4. Comparison of  $\hat{\sigma}$  and  $f_s$  across departments, journals, and researchers.** We show the maximum likelihood fitted  $\hat{\sigma}$  (**top**) and the fraction of secondary publications (**bottom**) for chemistry departments and chemistry journals in select years, and for all chemistry researchers in our database. The black horizontal dashed lines mark the value of the corresponding parameter for the *Journal of the American Chemical Society* in 1995. For clarity, we do not show  $\hat{\sigma}$  for 19 journals and 9 researchers that are outliers.

Parameter	Discipline	slope ( $m$ )	intercept ( $b$ )	$R^2$	$p$
$\hat{\mu} = mN_p + b$	ChemEng	0.051	1.218	0.00187	0.5240
	Chemistry	0.073	1.286	0.00668	0.0744
	Ecology	-0.150	1.658	0.00844	0.4502
	IndustEng	0.379	0.348	0.04305	0.3166
	MatScience	0.106	1.043	0.01114	0.1278
	MolBio	0.156	1.332	0.01909	0.0410
	Psychology	0.104	1.229	0.00496	0.5732
$\hat{\sigma} = mN_p + b$	ChemEng	0.067	0.379	0.03542	0.0052*
	Chemistry	0.031	0.418	0.00650	0.0862
	Ecology	0.059	0.434	0.00806	0.4524
	IndustEng	-0.094	0.764	0.01657	0.5380
	MatScience	0.033	0.502	0.00921	0.1598
	MolBio	0.064	0.415	0.01688	0.0592
	Psychology	-0.092	0.761	0.02026	0.2302

**Table S1. Individual lognormal parameters show no dependence on  $N_p$**  For each researcher within each of the seven disciplines we perform least-squares linear regression between the lognormal parameters  $\hat{\mu}$  and  $\hat{\sigma}$ , and  $\log_{10}(N_p)$ . We used a permutation test to calculate the  $p$ -values: for each set of pairs,  $(\hat{\mu}, N_p)$  and  $(\hat{\sigma}, N_p)$ , we performed 10,000 random swaps of all  $N_p$  and subsequent regression; we obtained a  $p$ -value by comparing the original slope of the fit with the distribution of the permuted slopes. \* $p < 0.05/7 \sim 0.0074$ .

Parameter	Discipline	Mean	Std Dev	Min	Median	Max
$\hat{\mu}$	ChemEng	1.354	0.256	0.566	1.320	1.921
	Chemistry	1.439	0.225	0.689	1.437	2.179
	Ecology	1.395	0.308	0.702	1.375	1.999
	IndustEng	1.012	0.313	0.603	1.046	1.466
	MatScience	1.250	0.266	0.629	1.254	1.947
	MolBio	1.624	0.253	0.950	1.641	2.250
	Psychology	1.437	0.318	0.545	1.427	1.967
$\hat{\sigma}$	ChemEng	0.508	0.080	0.323	0.513	0.764
	Chemistry	0.468	0.095	0.300	0.482	0.956
	Ecology	0.536	0.124	0.359	0.546	0.896
	IndustEng	0.604	0.150	0.438	0.591	0.796
	MatScience	0.559	0.095	0.335	0.568	0.969
	MolBio	0.525	0.105	0.362	0.541	1.006
	Psychology	0.550	0.137	0.398	0.585	0.954

Table S2. Individual discipline statistics of the lognormal model parameters